

Ana Paula de Oliveira

# **Le Traitement Automatique des Langues: vers une possible informatisation de l'enseignement ?**

**RÉSUMÉ:** La technologie n'a cessé de se développer au cours du siècle dernier, ce qui a conduit les enseignants et chercheurs à s'interroger de plus en plus sur la place du TAL au sein de nos salles de classe. Mais, comment exploiter favorablement ces ressources ? Partant de différents documents authentiques (articles de presse de la Première Guerre Mondiale) nous nous sommes attachés à construire une proposition d'enseignement des langues et de la linguistique à travers le prisme de l'informatique, de l'intelligence artificielle, mais également de l'histoire. Nous nous proposons, à travers une étude comparative de ces articles de presse, de mettre en place un enseignement de la langue en ASC (Apprentissage sur corpus) par le biais d'approches lexicométriques et textométriques réalisées au moyen de notre logiciel multiplateforme TXM. Cette analyse exhaustive et assistée par ordinateur nous offre une approche originale du corpus qui nous permettra, peut-être, d'aller vers une possible informatisation de l'enseignement.

**MOTS-CLÉS:** enseignement, apprentissage sur corpus, TXM, lexicométrie, textométrie

## **1. Le TAL au sein de l'enseignement**

À la fin du XXe siècle et au début du XXIe siècle, la fusion entre l'informatique et les textes a permis à la langue et à l'écriture de se réinventer et de se développer. En effet, d'innombrables textes se sont vus numérisés, digitalisés ou même tapés entièrement et manuellement, ce qui a permis, par extension, d'archiver et de consulter en un clic toute une série de documents et de ressources extrêmement précieuses pour la recherche et pour l'humanité toute entière.

### **1.1. Du TIC au TAL**

Ce nouvel espace d'apprentissage dit TIC, fortement implanté depuis une décennie dans les écoles ou centres éducatifs, nous conduit désormais, et de façon naturelle, vers le Traitement du Langage Naturel (TLN ou TAL) et les nombreuses possibilités qu'offre celui-ci. Il est, d'une certaine manière, le prolongement et l'approfondissement de ces nouvelles technologies ; une sorte de nouvelle cohabitation, nécessaire et inévitable, entre la linguistique et l'informatique. Rappelons la première définition du TAL donnée par Fuchs en 1993 :

[...] Les traitements automatiques des langues ont pour objet des données linguistiques (textes) exprimées dans une langue (naturelle), et (...) pour pouvoir traiter automatiquement ces données, il faut être capable d'explicitier les règles de la langue, de les représenter dans des formalismes opératoires et calculables et de les implémenter à l'aide de programmes (Fuchs, 1993, p. 13).

Signalons que, même si ces deux domaines se côtoient depuis très longtemps, cet engouement récent pour les nouvelles technologies et, notamment l'outil informatique, a suscité de nouvelles problématiques sur lesquelles s'interrogent différents spécialistes : « l'utilisation de l'informatique dans un cadre linguistique est très ancienne [...] L'utilisation massive de l'informatique semble créer, de fait, une modification du profil de certains acteurs de la recherche en linguistique : suis-je un informaticien ou un linguiste ? » (Issac, 2014, p. 1). En somme, cette dualité de deux entités contiguës ne cesse de faire surgir de nouveaux défis et enjeux.

Une question d'autant plus importante car elle façonne voire conditionne la pertinence des recherches et des résultats afin de garantir leur correcte interprétation. Le linguiste se voit désormais doté d'un outil extrêmement performant qui est celui de l'informatique, mais il ne doit en aucun cas s'éloigner de sa fonction première.

Ces outils apportent de nombreuses « pistes » de lecture au linguiste qu'il doit ensuite analyser scrupuleusement depuis son propre domaine scientifique. « Il faut sans doute avouer que l'informatique et le TAL peuvent aider à résoudre une bonne quantité des problèmes liés à la linguistique avec des taux de réussite assez significatifs, mais il faudra toujours le travail du linguiste afin de vérifier la production de la machine ». (Molina, 2014, p. 54).

Suite à cela, la question suivante surgit : le TAL est-il une science à la croisée entre l'informatique et la linguistique ? Et si tel est le cas, où le situer et comment définir son rôle au sein de l'enseignement ? Issac (2014) avance une réponse à ces questions en évoquant la fonction du linguiste informaticien « L'enjeu consiste donc à nommer et décrire de quelle manière l'informatique apporte de nouveaux outils et par là même de nouveaux regards théoriques sur la linguistique » (Issac, 2014, p. 13).

Il est évident que le TAL est un outil extrêmement performant pour un philologue qui se passionne à « disséquer » un texte afin de mieux l'interpréter et le comprendre car le TAL offre la possibilité d'analyses complexes et précises sur de multiples corpus extrêmement volumineux.

Dans un futur proche, il pourrait éventuellement devenir un outil indispensable pour aider le philologue et le professeur dans ses tâches quotidiennes d'analyse du discours car les résultats sont révélateurs, comme l'indique le

linguiste Robert Martin, « [...] l'analyse automatique contraint à la formalisation, c'est-à-dire à des formulations explicites et univoques [...] Soumise à l'automatisation, la description linguistique gagne en rigueur et parfois en pertinence » (Martin, 2001).

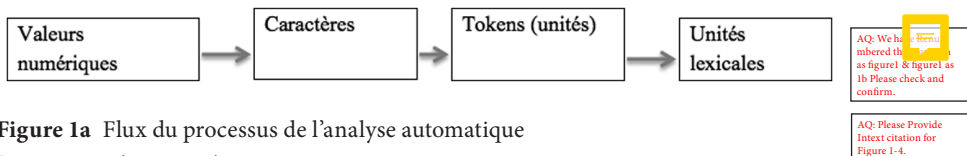
## 1.2. Humanités numériques

Le TAL est donc un instrument performant qui se définit comme « pluridisciplinaire puisqu'il peut faire collaborer linguistes, informaticiens, logiciens, psychologues, documentalistes, lexicologues et traducteurs ». En somme, pas moins de sept domaines en corrélation continue grâce à l'Intelligence Artificielle (IA) comme le signale l'Encyclopédie Universalis de France.

La linguistique mathématique semble se trouver aujourd'hui dans une position entièrement favorable, parmi les approches mathématiques en sciences sociales et physiologiques, pour se développer non seulement comme simple théorie des faits, mais aussi comme l'étude des principes et des structures hautement abstraites qui déterminent le caractère des processus mentaux humains. En ce cas, les processus mentaux en question sont ceux impliqués dans l'organisation d'un domaine spécifique de la connaissance humaine, à savoir : la connaissance de la langue (Chomsky, 1968, p. 107).

Au niveau scientifique, ce contexte d'informatisation des bibliothèques et d'automatisation de la recherche documentaire est relativement récent et est propre à notre siècle. La statistique textuelle permet désormais, entre autres, de reconnaître automatiquement les mots. Son processus s'organise en quatre étapes : le découpage de la chaîne, son analyse morphologique, syntaxique et enfin sémantique.

Nous pourrions schématiser ce processus informatiquement antérieurement verbalisé de la manière suivante :



Grâce à cette configuration au sein de notre machine, des liens hypertextes (catégories grammaticales, concordances, cooccurrences, etc.) peuvent être sollicités à travers une simple requête, ce qui est significatif pour le travail de recherche en science humaine.

### 1.3. Enjeux et perspectives

En ce qui concerne le domaine scientifique du TAL, cette nouvelle approche théorique représente une source intarissable de perspectives et d'interprétations pour le lecteur et lui propose d'autres parcours de lectures sur des corpus de texte, par l'observation de la fréquence et la disposition des mots, mais aussi et surtout par leur contextualisation.

Il s'agit là d'une linguistique de corpus au sens où la définit Teubert (2009, p. 1) : « la linguistique de corpus a été développée pour extraire d'un corpus les connaissances linguistiques nécessaires à l'enseignement des langues ; un corpus étant un ensemble collecté et ordonné de données langagières réelles ». Concernant les auteurs français, ils qualifient les corpus « de grandes collections textuelles monolingues ou bilingues qu'il s'agit d'étiqueter, d'analyser ou d'aligner » (Daille & Romary, 2001, p. 343).

En définitive, le TAL est une sorte de prise de contact synthétique du corpus qui permet de mettre en évidence les spécificités et les contrastes significatifs ou encore les phénomènes qui se répètent dans le discours. En effet, cette approche automatisée sur l'ensemble du corpus apporte des connaissances, des éléments curieux ou singuliers qui devront être interprétés avec précaution. Néanmoins, elle propose, d'emblée, un certain recul par rapport à l'ensemble du texte car elle en extrait des informations brutes et les positionne par ordre de fréquence selon les différentes requêtes sollicitées.

On parle alors d'un ensemble de résultats structurés qui peuvent représenter un facteur d'ouverture ou de nouvelles pistes de réflexion car, comme le souligne Pincemin<sup>1</sup>, il s'agit d'une « réalité langagière devenue un terrain d'observation important et très riche pour les sciences humaines et sociales ».

En somme, les fonctions d'exploration et de synthèse du TAL sont au service d'une vision globale, qualitative et quantitative proposant une pluralité de résultats qu'il faudra, par la suite, mettre en corrélation pour de possibles interprétations.

---

1 « Qu'est-ce que la textométrie ? Présentation », Site du projet Textométrie, <http://textometrie.ens-lyon.fr/?lang=fr>. Consulté le 04/07/2022.

## 2. Principales opérations informatiques

### 2.1. Importation de notre corpus

Notre plateforme logicielle est TXM car elle est open-source et combine des techniques puissantes et originales pour l'analyse d'opérations très variées et complexes sur des textes volumineux.

Sur l'interface de la machine nous disposons d'une option qui permet l'import de notre corpus grâce à la fonction « import ». Pour importer notre corpus, il sera nécessaire de convertir nos documents provenant du traitement de texte de type doc ou pdf au format txt afin qu'ils soient compatibles avec notre machine et lus correctement par celle-ci. L'import se fait toujours avec des documents en format txt ainsi que leur excel (CSV). Un fois que l'importation est correctement réalisée, une boîte représentant notre corpus apparaît sur la fenêtre de gauche.

Tous les résultats de calculs sous formes de tableaux sont exportables en CSV<sup>2</sup> et manipulables par d'autres logiciels. Les graphiques sont aussi exportables sous forme d'images vectorielles ce qui est très pratique et utile pour des supports de classe ou des publications.

TXM est développé avec les deux langages de programmation les plus utilisés de l'industrie informatique Java et C. Ce qui rend ce programme plus facile à utiliser car les utilisateurs sont déjà familiarisés avec ce premier langage. Cela le rend également plus compatible car nos machines disposent déjà de ce langage.

De plus, signalons que TXM peut s'installer sur toutes les plateformes (Windows, Mac OS X et Linux), mais qu'il est également disponible directement en ligne, sous forme de serveur web de portail, c'est-à-dire, sans aucune installation. Il suffit alors d'y importer son corpus.

### 2.2. Étiquetage morphosyntaxique

L'étiquetage morphosyntaxique est l'étape qui consiste à ajouter des propriétés morphosyntaxiques et des lemmes. Ce système d'automatisation de reconnaissance morphosyntaxique est basé sur la reconnaissance de par la position. Tout comme le ferait un linguiste, la machine est capable de déterminer la catégorie grammaticale grâce à la position du mot dans l'ensemble de la phrase.

Cet étiquetage automatisé permet d'associer à chaque unité textuelle des informations morphosyntaxiques (lemme, catégorie grammaticale, genre, nombre, temps, etc). En d'autres termes, c'est le fait d'explicitier les règles de la

---

2 Format Excel.

langue. Comme le définissent Vergne et Giguët (1998, p. 1) : « le « tagging », ou « étiquetage », ou « marquage », consiste à affecter une « étiquette » (« tag » ou catégorie) à chaque « mot » d'un texte ».

Grâce à cette tokenisation, c'est-à-dire à cet étiquetage morphosyntaxique lorsque l'on positionne le curseur sur le terme « Allemagne », par exemple, celui-ci nous indique qu'il s'agit d'un Nom Propre (catégorie NAM). De la même manière, si nous avons placé notre curseur sur un nom, un adjectif ou un autre terme nous aurions obtenu l'information correspondante.

TreeTagger est une sorte d'analogie au raisonnement du linguiste car la machine possède désormais une intelligence artificielle interne de reconnaissance capable d'étiqueter automatiquement des propriétés morphosyntaxiques et des lemmes aux mots de notre corpus. Nous obtenons ainsi, en retour, toute une série d'informations et de précisions concernant ce qui était, à la base, uniquement un ensemble de caractères. Sans TreeTagger, il serait donc impossible d'effectuer des calculs de reconnaissance morphosyntaxique.

### 2.3. Requêtes et exploitations

Nous lançons une première requête qui va nous permettre de calculer le nombre d'entrées selon la catégorie grammaticale sollicitée tout en respectant la distribution initiale de la partition. Pour ce faire, nous faisons clic droit sur la partition puis « calcul de la table lexicale » et nous choisissons l'option de « frpos ».

La Requête [frpos] sur la partition, nous permet d'identifier la fréquence de positionnement des différentes catégories grammaticales au sein de l'ensemble du corpus. Si nous lançons cette requête, nous obtenons les résultats suivants :

Ces premiers résultats nous ont servi, au cours de nos recherches, de socle pour le reste de notre étude car ils nous ont permis de prendre conscience instantanément des éléments qui ont été privilégiés dans ce discours journalistique et de dégager les fonctions principales qui l'articulent.

frpos v 32, Fmin 2, Fmax 13293

Appliquer Nombre de lignes 32 Fmin : 2

Fusion ou Suppr. de colonnes Fusion ou Suppr. de lignes

| frpos    | Fréquence | L'Humanite t=12734 | LeTemps t=23897 | LeFigaro t=16508 | LaCroix t=12536 |
|----------|-----------|--------------------|-----------------|------------------|-----------------|
| NOM      | 13293     | 2531               | 4987            | 3238             | 2537            |
| PRP      | 8514      | 1627               | 3175            | 2101             | 1611            |
| DET:ART  | 6872      | 1296               | 2480            | 1772             | 1324            |
| PUN      | 5035      | 970                | 1947            | 1296             | 822             |
| ADJ      | 3592      | 801                | 1258            | 860              | 673             |
| NAM      | 3477      | 792                | 1187            | 826              | 672             |
| VER:pres | 2857      | 525                | 934             | 752              | 646             |
| SENT     | 2820      | 603                | 970             | 688              | 559             |
| ADV      | 2437      | 443                | 847             | 641              | 506             |
| VER:ppre | 2379      | 403                | 845             | 628              | 503             |
| KON      | 2356      | 452                | 871             | 601              | 432             |
| PRO:PER  | 2187      | 406                | 736             | 620              | 425             |
| PRP:det  | 2132      | 459                | 744             | 498              | 431             |
| VER:infi | 1452      | 277                | 517             | 370              | 288             |
| NUM      | 1290      | 193                | 699             | 227              | 171             |
| PRO:REL  | 868       | 156                | 284             | 258              | 170             |
| DET:POS  | 788       | 151                | 269             | 198              | 170             |
| PRO:DEM  | 718       | 146                | 240             | 200              | 132             |
| PRO:IND  | 557       | 113                | 184             | 151              | 109             |
| VER:futu | 478       | 103                | 186             | 116              | 73              |
| VER:impf | 435       | 71                 | 155             | 135              | 74              |
| VER:ppre | 351       | 55                 | 133             | 93               | 70              |
| PUN:cit  | 218       | 30                 | 78              | 80               | 30              |
| VER:cond | 185       | 59                 | 54              | 43               | 29              |

Étiquetage morphosyntaxique – requête [frpos] dans la partition

**FIGURE 1b** Positionnement des différentes catégories grammaticales

Source : Extrait de notre corpus [partition regroupant les 4 lignes éditoriales].

Cette première observation nous a conduit à une réflexion principale : la nominalisation puisqu'il s'agit d'un procédé journalistique récurrent, ou encore la prédication nominale et l'analyse par verbes supports. Notre corpus est donc propice à des exercices portant sur les substantifs, mais il permet également de mettre en place différents exercices portant sur tous les éléments de morphosyntaxe.

Au sein de notre dernière partie, nous nous attachons donc à l'exploitation de notre corpus avec des activités en ASC (apprentissage sur corpus). Les différents points abordés seront le CECRL (Cadre Européen Commun de Référence pour les Langues), l'exploitation morphosyntaxique et lexicale à travers la mise en place d'exercices semi-automatisés que nous avons élaborés en nous basant sur les composantes de morphologie, de syntaxe et de lexique requises au niveau du CECRL.

Ces activités vont permettre d'acquérir des notions telles que la cohérence et la cohésion textuelles et discursives, les notions de figements, de semi-figements, de prédicats, mais aussi de GN, GV, de prépositions, d'accords, etc. C'est-à-dire

qu'on va pouvoir travailler sur morphologie, mais également la syntaxe grâce au cadre phrastique.

### **3. Exploitation du corpus à travers des activités en ASC**

#### **3.1. Qu'est-ce que l'ASC**

De nombreuses études menées dans le domaine de la didactique des langues ont montré les conséquences positives de l'ASC sur les apprenants. Comme le soulignent Boulton et Tyne (2014) qui citent, eux-mêmes, les résultats des travaux menés par Allan et Johns « Autrement dit, il semblerait qu'il y ait plus de chances que les apprenants ayant connu une expérience d'ASC dans une application directe deviennent, en quelque sorte, de meilleurs apprenants » (Boulton & Tyne, 2014, p. 162) car selon Kübler, il s'agit d'un support polyvalent et « on comprend la manière de communiquer dans une langue à partir du corpus, le corpus étant un réservoir inépuisable de découvertes, d'exemples et d'activités » (Kübler, 2014).

On s'est donc proposées de créer des séquences didactiques semi-automatisées. Néanmoins, nous proposerons, dans cet article, de simples échantillons car le but principal était surtout celui d'exposer la démarche de réflexion des exercices semi-automatisés.

#### **3.2. Activités en ASC – morphosyntaxe**

La requête du *\**  par exemple, nous permet de matcher, dans l'ensemble du corpus et en un seul clic tout ce qui est avant "ation" (pour identifier les noms qui proviennent de verbes, par exemple). Dans un autre exercice, on pourrait envisager l'exploitation des adjectifs *\*able*  (qui proviennent des verbes), etc.



a. *Trouvez les noms dérivés du verbe, de l'adjectif, puis du nom et indiquez leur genre :*

*Verbe -> nom : Formez les noms à partir des verbes proposées, puis continuer les listes.*

| Terminaisons | Requêtes                               |                           |
|--------------|----------------------------------------|---------------------------|
| -ation       | [word="*.ation"] <sup>1</sup>          | Mobiliser -> mobilisation |
| -ment        | [frpos="ADV"]<br>ou<br>[word="*.ment"] | Gouverner -> gouvernement |
| -able        | [word="*.able"]                        | Mobiliser -> mobilisable  |
| -ure         | [word="*.ure"]                         | Mesurer -> mesure         |
| -age         | [word="*.age"]                         | Passer -> passage         |
| -ée          | [word="*.ée"]                          | Armer -> armée            |
| -ance        | [word="*.ance"]                        | Confier -> confiance      |
| -eur         | [word="*.eur"]                         | Gouverner -> gouverneur   |
| -euse        | [word="*.euse"]                        | Danger -> dangereuse      |

FIGURE 2 Exemple de recherche sur corpus

Source : Extrait de notre corpus [partition regroupant les 4 lignes éditoriales]

On aurait également les requêtes avec les différentes constructions verbales suivies des prépositions « à » ou « de » :

- **Les constructions verbales**

Requêtes : [frpos="VER:pres"][word="à"], puis [frpos="VER:pres"][word="de"]

Complétez le texte avec les prépositions qui conviennent :

« Tous nos soldats répondent ..... l'appel de Dieu. L'Allemagne qui s'apprête ... envahir la Belgique s'étonne ... la force de notre nation. M. Schoen est le dernier Allemand qui reste ... Paris. On sait que l'un des plans allemands consiste ... foncer sur Paris. Le but est ... nous provoquer à la frontière même. La France demande ... l'aide ». (Le Temps)

Ou encore un exercice avec les accords au participe passé :

- **L'accord du participe passé**

Requête : [frlemma="avoir"][frpos="VER:pper"]

Observez :

Pauvres soldats, les balles qu'ils ont reçues, les ont anéantis.

Complétez en faisant l'accord, si nécessaire.

Les prêtres appelés ont reçu tous pouvoirs. Ils les ont reçu....

Ce discours a fait verser des larmes. Il les a fait versé....

Le gouvernement a pris des mesures. Il les a pris....

L'Allemagne a déclaré la guerre. Elle l'a déclaré....

Le président de la République a signé les décrets. Il les a signé....

Ou encore des exercices de constructions ou de cohérences verbales avec différents temps (passé composé, imparfait, passé simple, etc.)

#### **Passé composé, imparfait ou passé simple ?**

Requêtes :

[frpos="VER:pper"] (participe passé) ; [frpos="VER:impf"] (imparfait) ou [frpos="VER:simp"] (passé simple)

« L'Allemagne (déclarer) ..... la guerre à la Russie. De nombreux hommes (appeler) ..... sous les drapeaux. Tous les officiers qui (se trouver) ..... en ville (accourir) ..... au port et l'ordre de mobilisation (être) ..... affiché. Les passants (commenter) ..... les événements tous les matins et on (voir) ..... des gens se hâter ». (La Croix)

On propose ici des exercices à trous, mais le but est que les étudiants, eux-mêmes, puissent y importer leur corpus afin de développer une logique des constructions morphosyntaxiques.

### **3.3. Activités en ASC – lexique**

Concernant le lexique, on peut créer un inventaire du lexique, répertorié par domaine et en accord avec le CECRL. Par exemple, la requête [frpos="NOM"] nous permet, en un clic, d'obtenir une liste détaillée du lexique (substantif) présent dans notre corpus et d'en élaborer une fiche lexicale répertoriée par statistique lexicale :

| LEXIQUE      |              |              |              |                |
|--------------|--------------|--------------|--------------|----------------|
| Guerre       | Départements | Forces       | Puissances   | Parti          |
| Mobilisation | Pays         | Ministère    | Terre        | Pouvoirs       |
| Août         | Police       | Patrie       | Coup         | Actes          |
| Gouvernement | Président    | Réunion      | Esprit       | Autorités      |
| Ministre     | Sens         | Dispositions | Groupes      | Classes        |
| Armée        | Droit        | Confiance    | Infanterie   | Cours          |
| Hommes       | Autorité     | Préfet       | Journaux     | Emotion        |
| Service      | Ministres    | Services     | Traité       | Dépêche        |
| Paix         | Ambassadeur  | Situation    | Citoyens     | Membres        |
| Ordre        | Mesures      | Défense      | Armes        | télegrammes    |
| Etat         | Régiments    | Engagement   | Bataillons   | Dieu           |
| Frontière    | Devoir       | Conditions   | Cavalerie    | Gens           |
| Siège        | Population   | Public       | Etrangers    | Manifestations |
| Territoire   | Drapeaux     | Régiment     | Prix         | Note           |
| Décret       | Monde        | Soldats      | Art          | Réserve        |
| Déclaration  | Ville        | Raison       | Conseil      | Volontaires    |
| Troupes      | Foule        | Attitude     | Dignité      | Article        |
| Loi          | Officiers    | Avis         | Efforts      | Calme          |
| Neutralité   | Division     | Mesure       | Enthousiasme | Chef           |
| République   | Événements   | Patriotisme  | Familles     | Dévouement     |

FIGURE 3 Fiche lexicale par statistiques lexicales

Source : Extrait de notre corpus [partition regroupant les 4 lignes éditoriales]

À partir de cette fiche lexicale, nous pouvons créer également des exercices portant sur les champs lexicaux, les champs sémantiques, etc. ou même des fiches lexicales répertoriées par thématique :

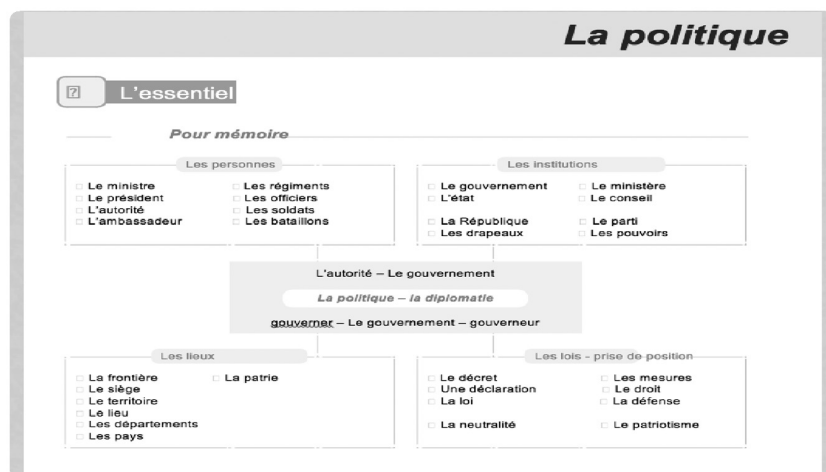


FIGURE 4 Fiche lexicale par thématique

Source : Extrait de notre corpus [partition regroupant les 4 lignes éditoriales]

## 4. Conclusion

En guise de conclusion et comme perspectives futures, nous estimons que le TAL devrait pouvoir s'imposer comme un outil de plus au sein de l'enseignement, car il propose une exploitation exhaustive, rapide et précise des corpus, ainsi que la possibilité de conception d'exercices semi-automatisés, ce qui est un atout majeur pour les professeurs de FLE.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Boulton, A., & Tyne, H. (2014). *Des documents authentiques aux corpus*. Éditions Didier.
- Chomsky, N. (1968). *Le langage et la pensée*. Petite bibliothèque Payot.
- Daille, B., & Romary, L. (2001). Linguistique de corpus. Éditorial. *Traitement automatique des langues*, 42(2), 343-346.
- Fuchs, C. (1993). *Linguistique et traitements automatiques des langues*. Éditions Hachette Université Linguistique. Hachette Supérieur.
- Issac, F. (2014). *Du caractère au dictionnaire, méthodes et outils pour l'élaboration de ressources linguistiques* [Habilitation à diriger des recherches]. Université Paris 13 Sorbonne.

- Kübler, N. (2014). Mettre en œuvre la linguistique de corpus à l'université : Vers une compétence utile pour l'enseignement/apprentissage des langues ? Recherches en didactique des langues et des cultures. *Les Cahiers de l'Acedle*, 11(1). <https://doi.org/10.4000/rdlc.1685>
- Martin, R. (2001). *Sémantique et automate*. Éditions PUF.
- Molina Mejia, J-M. (2014). *ELITE-[FLE] : Un environnement d'ALAO fondé sur la linguistique textuelle, pour la formation linguistique des futurs enseignants de FLE en Colombie* [Thèse de doctorat, Université Grenoble Alpes]. HAL thèses. <https://tel.archives-ouvertes.fr/tel-01228439>
- Teubert, W. (2009). La linguistique de corpus : une alternative [version abrégée]. *SEMEN. Revue de Sémio-linguistique des Textes et Discours*, 27, 185–211. <https://doi.org/10.4000/semen.8914>
- Vergne, J., & Giguët, E. (1998). Regards théoriques sur le Tagging. In *Actes de la cinquième conférence Le Traitement Automatique des Langues Naturelles (TALN98)*. Association pour le Traitement Automatique des Langues.

